

Case 1. Energy consumption and price forecasting, 1-day ahead hourly

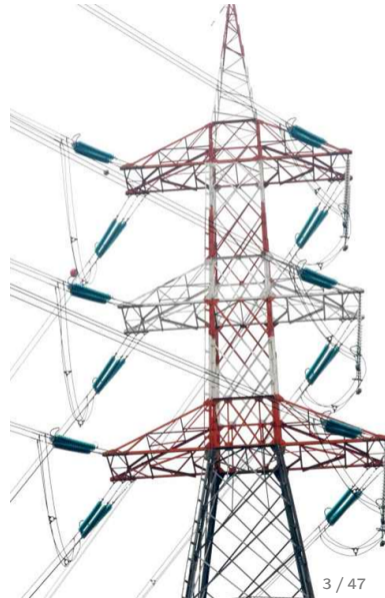
The components of multivariate time series with periodicity

Time series:

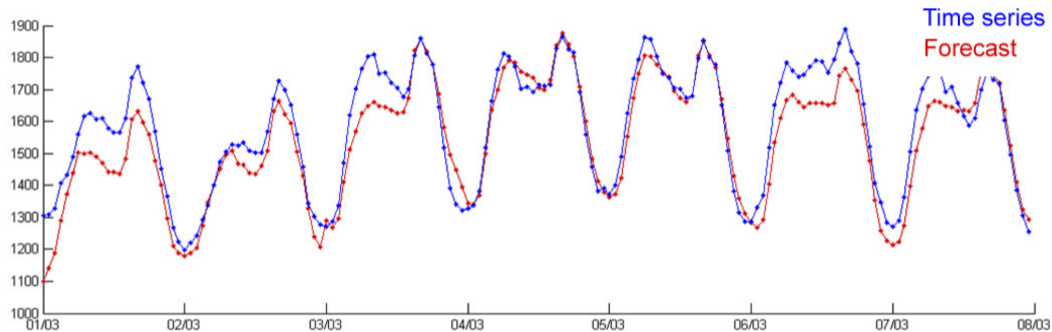
- ▶ energy price,
- ▶ consumption,
- ▶ daytime,
- ▶ temperature,
- ▶ humidity,
- ▶ wind force,
- ▶ holiday schedule.

Periodicity:

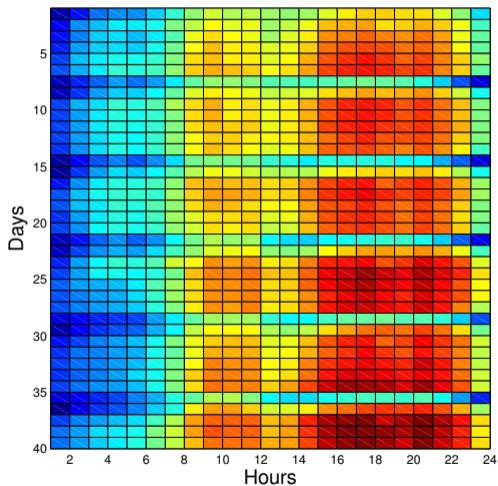
- ▶ one year seasons (temperature, daytime),
- ▶ one week,
- ▶ one day (working day, week-end),
- ▶ a holiday,
- ▶ aperiodic events.



Energy consumption one-week forecast for each hour



The autoregressive matrix, five weeks



The autoregressive matrix and the linear model

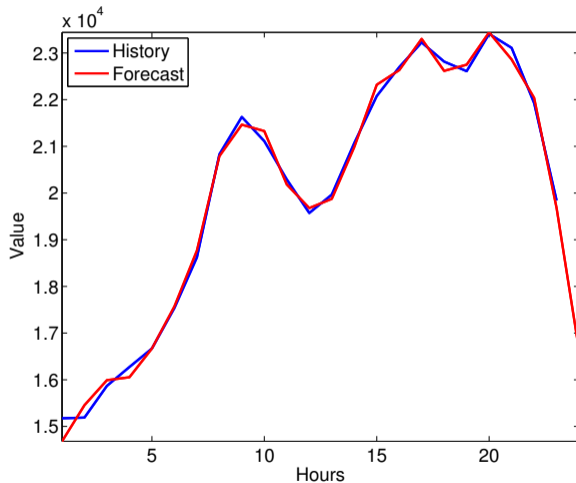
$$\mathbf{X}^*_{(m+1) \times (n+1)} = \left[\begin{array}{c|ccc} \hat{S}_T & S_{T-1} & \dots & S_{T-\kappa+1} \\ \hline S_{(m-1)\kappa} & S_{(m-1)\kappa-1} & \dots & S_{(m-2)\kappa+1} \\ \dots & \dots & \dots & \dots \\ S_{n\kappa} & S_{n\kappa-1} & \dots & S_{n(\kappa-1)+1} \\ \dots & \dots & \dots & \dots \\ S_\kappa & S_{\kappa-1} & \dots & S_1 \end{array} \right] = \left[\begin{array}{c|c} \hat{S}_T & \mathbf{x}_{m+1} \\ \hline \mathbf{y} & \mathbf{X} \end{array} \right].$$

The matrix \mathbf{X}^* is partitioned into a top row and a bottom block. The top row contains \hat{S}_T followed by $S_{T-1}, \dots, S_{T-\kappa+1}$. The bottom block consists of m rows, each containing a value from the S sequence followed by $\kappa-1$ lagged values. The right-hand side shows this as a linear model where the top row is a scalar \hat{S}_T and the bottom block is a vector \mathbf{y} of size $m \times 1$ and a matrix \mathbf{X} of size $m \times n$.

In terms of linear regression:

$$\hat{\mathbf{y}} = \mathbf{f}(\mathbf{X}, \mathbf{w}) = \mathbf{X}\mathbf{w},$$
$$\hat{y}_{m+1} = \hat{S}_T = \langle \mathbf{x}_{m+1}, \hat{\mathbf{w}} \rangle.$$

The one-day forecast: expected error is 3.1% working day, 3.7% week-end



The model $\hat{y} = f(\mathbf{X}, \mathbf{w})$ could be a linear model, neural network, deep NN, SVN, ...

The model performance criteria and forecast errors

Stability:

- ▶ the error does not change significantly following small changes in time series,
- ▶ the distribution of the model parameters does not change.

Complexity:

- ▶ the number of parameters (elements in superposition) is minimal,
- ▶ the minimum description length principle holds the William Ockham's rule.

Error: the residue $\varepsilon_j = \hat{y}_j - y_j$ for

- ▶ mean absolute error and (symmetric) mean absolute percent error

$$RSS = \sum_{j=1}^r \varepsilon_j^2, \quad MAPE = \frac{1}{r} \sum_{j=1}^r \frac{|\varepsilon_j|}{|y_j|}, \quad sMAPE = \frac{1}{r} \sum_{j=1}^r \frac{2|\varepsilon_j|}{|\hat{y}_j + y_j|}.$$